Package 'GWAF'

March 12, 2015

Type Package

Title Genome-Wide Association/Interaction Analysis and Rare Variant Analysis with Family Data

Version 2.2

Date 2015-03-12

Author Ming-Huei Chen <mhchen@bu.edu> and Qiong Yang <qyang@bu.edu>

Maintainer Ming-Huei Chen <mhchen@bu.edu>

Depends R(>= 2.15.1), geepack, coxme, lme4

Description

Functions for genome-wide association/interaction analysis and rare variant analysis on a continuous/dichotomous trait using family data, and for making genome-wide p-value plot and QQ plot.

License GPL (>= 2)

NeedsCompilation no

Repository CRAN

Date/Publication 2015-03-12 20:46:12

R topics documented:

GWAF-package *Genome-wide association/interaction analysis and rare variant analysis with family data*

Description

For continuous traits, GWAF package provides two sets of functions for each of genome-wide association/interaction analyses with observed/imputed SNP genotypes for family data. One fits Linear Mixed Effects (LME) model and the other fits Generalized Estimation Equation (GEE) model to accounting for within pedigree correlation. While for dichotomous trait, GWAF package provides functions to fit GEE model for genome-wide association/interaction analyses. For rare variant analysis, GWAF fits LME and generalized linear mixed effects (GLMM) model for continuous traits and dichotomous traits, respectively. In addition, GWAF package also provides functions for making genome-wide p-values plot and QQ plot that contains genomic control parameter estimate and generating scripts for genome-wide association analysis. Please read UsingGWAF.pdf for more information including examples and description to output files.

Details

Package:	GWAF
Type:	Package
Version:	2.1
Date:	2013-06-06
License:	GPL (>= 2)

Author(s)

Qiong Yang <qyang@bu.edu> and Ming-Huei Chen <mhchen@bu.edu>

Maintainer: Ming-Huei Chen <mhchen@bu.edu>

auto *function to generate scripts for genome-wide association/interaction analysis*

Description

Given a path/directory (genopath) that keeps geotype files, phenotype file, pedigree file, phenotype of interest, covariates, analysis of interest (can be 'lmepack', 'lmepack.imputed', 'lmeVpack.imputed', 'glmm', 'geepack', 'geepack.imputed', 'geepack.quant', 'geepack.quant.imputed', 'lmepack.int', 'lmepack.int.imputed', 'geepack.int', 'geepack.int.imputed', 'geepack.quant.int', 'geepack.quant.int.imputed') and other arguments, auto function generates one R script, one shell script that can excute R script, and one list file that can excute all shell scripts in batch mode, for each genotype file. Once the list file (XXXX.lst) is generated, user can use ksh XXXX.lst to submit all jobs to test all SNPs in genopath.

Usage

```
auto(genopath, phenfile, pedfile, outfile, phen, covars, cov.int, sub="N",
analysis, lib.loc, model = NULL, kinmat = NULL, col.names = F, sep.ped = ",",
sep.phe = ",", sep.gen = ",")
```

Arguments

genopath	a character string indicating the path/directory that keeps genotype files to be analyzed
phenfile	a character string naming the phenotype file for reading (see format requirement in details)
pedfile	a character string naming the pedigree file for reading (see format requirement in details)
outfile	a character string naming the result file for writing
phen	a character string for a phenotype name in phenfile
covars	a character vector for covariates in phenfile
cov.int	a character string naming the covariate for interaction, the covariate has to be included in covars
sub	"N" (default) for no stratified analysis, and "Y" for requesting stratified analyses (only when cov.int is dichotomous)
analysis	a character string indicating the analysis of interest available in GWAF package, can be 'lme', 'lme.imputed', 'gee' or 'gee.imputed'
lib.loc	a character string indicating the location of GWAF package
model	a single character of 'a','d','g', or 'r', with 'a'=additive, 'd'=dominant, 'g'=general and 'r'=recessive models; Not appropriate/needed for analyzing imputed SNPs
kinmat	a character string naming the file where kinship coefficient matrix is kept; needed for LME analyses

col.names	a logical value indicating whether the output file should contain column names
sep.ped	the field separator character for pedigree file
sep.phe	the field separator character for phenotype file
sep.gen	the field separator character for genotype file

Details

auto function generates one R script, one shell script that can excute R script, and one list file that can excute all shell scripts in batch mode. These scripts are named based on the phenotype of interest, the analysis of interest and the time these scripts are generated. After generating these scripts, auto function genertates a message telling the user how to submit ALL the jobs (using ksh XXXX.lst). When a submitted job is completed, a log file indicating which genotype file was analyzed will be generated and the R script and the shell script will be removed. The number of log files should equal to the number of genotype files, if all jobs are completed. All the results will be written and appended to the user specified single output file. Different outfile should be assigned for different genopath to avoid over-writting.

Value

No value is returned. Instead, results are written to outfile.

Author(s)

Ming-Huei Chen <mhchen@bu.edu> and Qiong Yang <qyang@bu.edu>

Examples

```
## Not run:
auto(phenfile="simphen.csv",genopath="/home/data/exomechip/chr1",pedfile="simped.csv",
outfile="exomechip_chr1_SIMQT.csv",phen="CVD",covars="sex",analysis="geepack",model="a",
col.names=F,sep.ped=",",sep.phe=",",sep.gen=",")

## End(Not run)
```

---

geepack.lgst *function for testing association between a dichotomous trait and a genotyped SNP in family data using Generalized Estimation Equation model*

Description

Fit logistic regression via GEE to test association between a dichotomous phenotype and one genotyped SNP in a genotype file with user specified genetic model. Each family is treated as a cluster, with independence working correlation matrix used in the robust variance estimator. The trait-SNP association test is carried out by the geese function from package geepack. This function is called in geepack.lgst.batch function to apply association test to all SNPs in the genotype data.

Usage

```
geepack.lgst(snp, phen, test.dat, covar = NULL, model = "a")
```

Arguments

snp	genotype data of a SNP
phen	a character string for a phenotype name in test.dat
test.dat	the product of merging phenotype, genotype and pedigree data, should be ordered by "famid"
covar	a character vector for covariates in test.dat
model	a single character of 'a','d','g', or 'r', with 'a'=additive, 'd'=dominant, 'g'=general and 'r'=recessive models

Details

The geepack.lgst function tests association between a dichtomous trait and a SNP from a dataset that contains phenotype, genotype and pedigree data (test.dat), where the dataset needs to be ordered by famid.

Value

Please see output in geepack.lgst.batch.

Author(s)

Qiong Yang <qyang@bu.edu> and Ming-Huei Chen <mhchen@bu.edu>

References

Liang, K.Y. and Zeger, S.L. (1986) Longitudinal data analysis using generalized linear models. *Biometrika*, **73** 13–22.

Zeger, S.L. and Liang, K.Y. (1986) Longitudinal data analysis for discrete and continuous outcomes. *Biometrics*, **42** 121–130.

Yan, J and Fine, J. (2004) Estimating equations for association structures. *Stat Med*, **23** 859–874.

See Also

geese function from package geepack

Examples

```
## Not run:
geepack.lgst(snp=data[,"rs123"],phen="CVD",test.dat=data,model="a",covar=c("age","sex"))

## End(Not run)
```

`geepack.lgst.batch` *function to test genetic associations between a dichotomous trait and a batch of genotyped SNPs in families using Generalized Estimation Equation model*

Description

Fit logistic regression via Generalized Estimation Equation (GEE) to test associations between a dichotomous phenotype and all genotyped SNPs in a genotype file in family data with user specified genetic model. Each pedigree is treated as a cluster, with independence working correlation matrix used in the robust variance estimator. This function applies the same trait-SNP association test to all SNPs in the genotype data. The trait-SNP association test is carried out by `geepack.lgst` function where the the `geese` function from package `geepack` is used.

Usage

```
geepack.lgst.batch(genfile, phenfile, pedfile, outfile, phen, covars = NULL,
model = "a", col.names = T, sep.ped = ",", sep.phe = ",", sep.gen = ",")
```

Arguments

`genfile`	a character string naming the genotype file for reading (see format requirement in details)
`phenfile`	a character string naming the phenotype file for reading (see format requirement in details)
`pedfile`	a character string naming the pedigree file for reading (see format requirement in details)
`outfile`	a character string naming the result file for writing
`phen`	a character string for a phenotype name in `phenfile`
`covars`	a character vector for covariates in `phenfile`
`model`	a single character of 'a','d','g', or 'r', with 'a'=additive, 'd'=dominant, 'g'=general and 'r'=recessive models
`col.names`	a logical value indicating whether the output file should contain column names
`sep.ped`	the field separator character for pedigree file
`sep.phe`	the field separator character for phenotype file
`sep.gen`	the field separator character for genotype file

Details

The `geepack.lgst.batch` function first reads in and merges phenotype-covariates, genotype and pedigree files, then tests the association of `phen` against all SNPs in `genfile`. `genfile` contains unique individual id and genotype data, with the column names being "id" and SNP names. For each genotyped SNP, the genotype data should be coded as 0, 1, 2 indicating the numbers of the coded

alleles. The SNP names in genotype file should not have any dash, '-' and other special characters(dots and underscores are OK). `phenfile` contains unique individual id, phenotype and covariates data, with the column names being "id" and phenotype and covaraite names. `pedfile` contains pedigree informaion, with the column names being "famid","id","fa","mo","sex". In all files, missing value should be an empty space, except missing parental id in `pedfile`. Only phenotypes with two categories are analyzed. A phenotype should be coded as 0 and 1, with 1 denoting affected and 0 unaffected. SNPs with low genotype counts (especially minor allele homozygote) may be omitted or analyzed with dominant model or analyzed with logistic regression. The `geepack.lgst.batch` function fits GEE model using each pedigree as a cluster with `geepack.lgst` function from GWAF package and `geese` function from `geepack` package.

Value

No value is returned. Instead, results are written to `outfile`. When the genetic model is 'a', 'd' or 'r', the result includes the following columns. When the genetic model is 'g', `beta` and `se` are replaced with `beta10`, `beta20`, `beta21`, `se10`, `se20`, and `se21` .

`phen`	phenotype name
`snp`	SNP name
`n0`	the number of individuals with 0 copy of coded alleles
`n1`	the number of individuals with 1 copy of coded alleles
`n2`	the number of individuals with 2 copies of coded alleles
`nd0`	the number of individuals with 0 copy of coded alleles in affected sample
`nd1`	the number of individuals with 1 copy of coded alleles in affected sample
`nd2`	the number of individuals with 2 copies of coded alleles in affected sample
`miss.0`	Genotype missing rate in unaffected sample
`miss.1`	Genotype missing rate in affected sample
`miss.diff.p`	P-value of differential missingness test between unaffected and affected samples
`beta`	regression coefficient of SNP covariate
`se`	standard error of `beta`
`chisq`	Chi-square statistic for testing `beta` not equal to zero
`df`	degree of freedom of the Chi-square statistic
`model`	model actually used in the analysis
`remark`	warning or additional information for the analysis, 'not converged' indicates the GEE analysis did not converge; 'logistic reg' indicates GEE model is replaced by logistic regression; 'exp count<5' indicates any expected count is less than 5 in phenotype-genotype table; 'not converged and exp count<5', 'logistic reg & exp count<5' are noted similarly; 'collinearity' indicates collinearity exists between SNP and some covariates
`pval`	p-value of the chi-square statistic
`beta10`	regression coefficient of genotype with 1 copy of coded allele vs. that with 0 copy

beta20	regression coefficient of genotype with 2 copy of coded allele vs. that with 0 copy
beta21	regression coefficient of genotype with 2 copy of coded allele vs. that with 1 copy
se10	standard error of beta10
se20	standard error of beta20
se21	standard error of beta21

Author(s)

Qiong Yang <qyang@bu.edu> and Ming-Huei Chen <mhchen@bu.edu>

Examples

```
## Not run:
geepack.lgst.batch(phenfile="simphen.csv",genfile="simgen.csv",pedfile="simped.csv",
phen="SIMQT",model="a",outfile="simout.csv",sep.ped=",",sep.phe=",",sep.gen=",")

## End(Not run)
```

---

geepack.lgst.batch.imputed

function to test genetic associations between a dichotomous trait and a batch of imputed SNPs in families using Generalized Estimation Equation model

---

Description

Fit logistic regression via Generalized Estimation Equation (GEE) to test associations between a dichotomous phenotype and all imputed SNPs in a genotype file in family data under additive genetic model. Each family is treated as a cluster, with independence working correlation matrix used in the robust variance estimator. This function applies the same trait-SNP association test to all SNPs in the imputed genotype data. The trait-SNP association test is carried out by geepack.lgst.imputed function where the geese function from package geepack is used.

Usage

```
geepack.lgst.batch.imputed(genfile, phenfile, pedfile, outfile, phen,
covars = NULL, col.names = T, sep.ped = ",", sep.phe = ",", sep.gen = ",")
```

Arguments

genfile	a character string naming the genotype file for reading (see format requirement in details)
phenfile	a character string naming the phenotype file for reading (see format requirement in details)

pedfile	a character string naming the pedigree file for reading (see format requirement in details)
outfile	a character string naming the result file for writing
phen	a character string for a phenotype name in phenfile
covars	a character vector for covariates in phenfile
col.names	a logical value indicating whether the output file should contain column names
sep.ped	the field separator character for pedigree file
sep.phe	the field separator character for phenotype file
sep.gen	the field separator character for genotype file

Details

Similar to the details for geepack.lgst.batch but here the SNP data contains imputed genotypes (allele dosages) that are continuous and range from 0 to 2. In addition, the user-specified genetic model argument is not available.

Value

No value is returned. Instead, results are written to outfile.

phen	phenotype name
snp	SNP name
N	the number of individuals in analysis
Nd	the number of individuals in affected sample in analysis
AF	imputed allele frequency of coded allele
AFd	imputed allele frequency of coded allele in affected sample
beta	regression coefficient of SNP covariate
se	standard error of beta
remark	warning or additional information for the analysis, note that the genotype counts are based on rounded imputed genotypes; 'not converged' indicates the GEE analysis did not converge; 'logistic reg' indicates GEE model is replaced by logistic regression; 'exp count<5' indicates any expected count is less than 5 in phenotype-genotype table; 'not converged and exp count<5', 'logistic reg & exp count<5' are noted similarly; 'collinearity' indicates collinearity exists between SNP and some covariates
pval	p-value of the association test based on chi-square statistic

Author(s)

Qiong Yang <qyang@bu.edu> and Ming-Huei Chen <mhchen@bu.edu>

Examples

```
## Not run:
geepack.lgst.batch.imputed(phenfile="simphen.csv",genfile="simgen.csv",
pedfile="simped.csv",phen="aff",covars="sex",outfile="simout.csv",col.names=T,
sep.ped=",",sep.phe=",",sep.gen=",")

## End(Not run)
```

`geepack.lgst.imputed` *function for testing association between a dichotomous trait and an imputed SNP in family data using Generalized Estimation Equation model*

Description

Fit logistic regression via Generalized Estimation Equation (GEE) to test association between a dichotomous phenotype and one imputed SNP in a genotype file in family data under additive genetic model. Each family is treated as a cluster, with independence working correlation matrix used in the robust variance estimator. The trait-SNP association test is carried out by the `geese` function from package `geepack`. This function is called in `geepack.lgst.batch.imputed` function to apply association test to all imputed SNPs in a genotype file.

Usage

```
geepack.lgst.imputed(snp, phen, test.dat, covar = NULL)
```

Arguments

`snp`	imputed genotype data of a SNP
`phen`	a character string for a phenotype name in `test.dat`
`test.dat`	the product of merging phenotype, genotype and pedigree data, should be ordered by "famid"
`covar`	a character vector for covariates in `test.dat`

Details

Similar to the details for `geepack.lgst` function but here the SNP data contains imputed genotypes (allele dosages) that are continuous and range from 0 to 2. In addition, the user-specified genetic model argument is not available.

Value

Please see output in `geepack.lgst.batch.imputed`.

Author(s)

Qiong Yang <qyang@bu.edu> and Ming-Huei Chen <mhchen@bu.edu>

References

Liang, K.Y. and Zeger, S.L. (1986) Longitudinal data analysis using generalized linear models. *Biometrika*, **73** 13–22.

Zeger, S.L. and Liang, K.Y. (1986) Longitudinal data analysis for discrete and continuous outcomes. *Biometrics*, **42** 121–130.

Yan, J and Fine, J. (2004) Estimating equations for association structures. *Stat Med*, **23** 859–874.

See Also

`geese` function from package `geepack`

Examples

```
## Not run:
geepack.lgst.imputed(snp=data[,"rs123"],phen="CVD",test.dat=data,covar=c("age",sex"))

## End(Not run)
```

`geepack.lgst.int` *function for testing gene-environment or gene-gene interaction between a dichotomous trait and a genotyped SNP in family data using Generalized Estimation Equation model*

Description

Fit logistic regression via Generalized Estimation Equation (GEE) to test gene-environment or gene-gene interaction between a dichotomous phenotype and one genotyped SNP in a genotype file in family data under additive genetic model. The interaction term is the product of SNP genotype and the covariate for interaction (`cov.int`). The covariate for interaction (`cov.int`) can be SNP genotype (gene-gene interaction) or an environmental factor (gene-environment interaction). Only one interaction term is allowed. When `cov.int` is dichotomous, stratified analyses can be requested by specifying `sub="Y"`. The covariance between the main effect (SNP) and the interaction effect is provided in the output when stratified analysis is not requested. Each family is treated as a cluster, with independence working correlation matrix used in the robust variance estimator. The interaction test is carried out by the `geese` function from package `geepack`. This function is called in `geepack.lgst.int.batch` function to apply interaction test to all SNPs in a genotype file.

Usage

```
geepack.lgst.int(snp,phen,test.dat,covar,cov.int,sub="N")
```

Arguments

`snp`	genotype data of a SNP
`phen`	a character string for a phenotype name in `test.dat`

test.dat	the product of merging phenotype, genotype and pedigree data, should be ordered by "famid"
covar	a character vector for covariates in test.dat
cov.int	a character string naming the covariate for interaction, the covariate has to be included in covar
sub	"N" (default) for no stratified analysis, and "Y" for requesting stratified analyses (only when cov.int is dichotomous)

Details

The geepack.lgst.int function tests gene-environment or gene-genn interaction between a dichtomous trait and a SNP from a dataset that contains phenotype, genotype and pedigree data (test.dat), where the dataset needs to be ordered by famid. Please also see details in details for geepack.lgst.int.batch function.

Value

Please see value in geepack.lgst.int.batch function.

Author(s)

Qiong Yang <qyang@bu.edu> and Ming-Huei Chen <mhchen@bu.edu>

References

Liang, K.Y. and Zeger, S.L. (1986) Longitudinal data analysis using generalized linear models. *Biometrika*, **73** 13–22.

Zeger, S.L. and Liang, K.Y. (1986) Longitudinal data analysis for discrete and continuous outcomes. *Biometrics*, **42** 121–130.

Yan, J and Fine, J. (2004) Estimating equations for association structures. *Stat Med*, **23** 859–874.

See Also

geese function from package geepack

Examples

```
## Not run:
geepack.lgst.int(snp=data[,"rs123"],phen="CVD",test.dat=data,covar=c("age",sex"),
cov.int="sex",sub="Y")

## End(Not run)
```

`geepack.lgst.int.batch` *function to test gene-environment or gene-gene interactions between a dichotomous trait and a batch of genotyped SNPs in families using Generalized Estimation Equation model*

Description

Fit logistic regression via Generalized Estimation Equation (GEE) to test gene-environment or gene-gene interactions for a dichotomous phenotype and all genotyped SNPs in a genotype file in family data under additive genetic model. The interaction term is the product of SNP genotype and a covariate for interaction (`cov.int`). The covariate for interaction (`cov.int`) can be SNP genotype (gene-gene interaction) or an environmental factor (gene-environment interaction). Only one interaction term is allowed. When `cov.int` is dichotomous, stratified analyses can be requested by specifying `sub="Y"`. The covariance between the main effect (SNP) and the interaction effect is provided in the output when stratified analysis is not requested. Each pedigree is treated as a cluster with independence working correlation matrix used in the robust variance estimator. This function applies the same interaction test to all SNPs in a genotype file. The interaction test is carried out by `geepack.lgst.int` function from `GWAF` where the the `geese` function from package `geepack` is used.

Usage

```
geepack.lgst.int.batch(genfile,phenfile,pedfile,outfile,phen,covars,cov.int,sub="N",
col.names=T,sep.ped=",",sep.phe=",",sep.gen=",")
```

Arguments

`genfile`	a character string naming the genotype file for reading (see format requirement in details)
`phenfile`	a character string naming the phenotype file for reading (see format requirement in details)
`pedfile`	a character string naming the pedigree file for reading (see format requirement in details)
`outfile`	a character string naming the result file for writing
`phen`	a character string for a phenotype name in `phenfile`
`covars`	a character vector for covariates in `phenfile`
`cov.int`	a character string naming the covariate for interaction, the covariate has to be included in `covars`
`sub`	"N" (default) for no stratified analysis, and "Y" for requesting stratified analyses (only when `cov.int` is dichotomous)
`col.names`	a logical value indicating whether the output file should contain column names
`sep.ped`	the field separator character for pedigree file
`sep.phe`	the field separator character for phenotype file
`sep.gen`	the field separator character for genotype file

Details

The geepack.lgst.int.batch function first reads in and merges phenotype-covariates, genotype and pedigree files, then tests gene-environment or gene-gen interaction for phen against all SNPs in genfile. Only one interaction term is allowed, so is the covariate for interaction (cov.int). When cov.int is dichotomous, stratified analyses can be requested by specifying sub="Y". The covariance between the main effect (SNP) and the interaction effect is provided in the output when stratified analysis is not requested. genfile contains unique individual id and genotype data, with the column names being "id" and SNP names. For each genotyped SNP, the genotype data should be coded as 0, 1, 2 indicating the numbers of the coded alleles. The SNP names in genotype file should not have any dash, '-' and other special characters(dots and underscores are OK). phenfile contains unique individual id, phenotype and covariates data, with the column names being "id" and phenotype and covaraite names. pedfile contains pedigree informaion, with the column names being "famid","id","fa","mo","sex". In all files, missing value should be an empty space, except missing parental id in pedfile. Only phenotypes with two categories are analyzed. A phenotype should be coded as 0 and 1, with 1 denoting affected and 0 unaffected. SNPs with low genotype counts (especially minor allele homozygote) may be omitted or analyzed with logistic regression. The geepack.lgst.int.batch function fits GEE model using each pedigree as a cluster with geepack.lgst.int function from GWAF package and geese function from geepack package.

Value

No value is returned. Instead, results are written to outfile. If stratified analyses are requested, the result file will include the following columns. Otherwise, cov_beta_snp_beta_int will be included instead of the results from stratified analyses, that is, beta_snp_cov0, se_snp_cov0, pval_snp_cov0, beta_snp_cov1, se_snp_cov1, and pval_snp_cov1.

phen	phenotype name
snp	SNP name
covar_int	the covariate for interaction
n	sample size used in analysis
AF	allele frequency of the coded allele
nd	the number of individuals in affected sample
AFd	allele frequency of the coded allele in affected sample
model	genetic model used in analysis, additive model only
beta_snp	regression coefficient of SNP covariate
se_snp	standard error of beta_snp
pval_snp	p-value of testing beta_snp not equal to zero
beta_snp_cov0	regression coefficient of SNP covariate in stratified analysis using the subset where cov.int level is 0
se_snp_cov0	standard error of beta_snp_cov0
pval_snp_cov0	p-value of testing beta_snp_cov0 not equal to zero
beta_snp_cov1	regression coefficient of SNP covariate in stratified analysis using the subset where cov.int level is 1
se_snp_cov1	standard error of beta_snp_cov1

pval_snp_cov1	p-value of testing beta_snp_cov1 not equal to zero
beta_int	regression coefficient of the interaction term
se_int	standard error of beta_int
pval_int	p-value of testing beta_int not equal to zero
remark	warning or additional information for the analysis, 'not converged' indicates the GEE analysis did not converge; 'logistic reg' indicates GEE model is replaced by logistic regression; 'exp count<5' indicates any expected count is less than 5 in phenotype-genotype table; 'not converged and exp count<5', 'logistic reg & exp count<5' are noted similarly; 'collinearity' indicates collinearity exists between SNP and some covariates

Author(s)

Qiong Yang <qyang@bu.edu> and Ming-Huei Chen <mhchen@bu.edu>

Examples

```
## Not run:
geepack.lgst.int.batch(phenfile="simphen.csv",genfile="simgen.csv",pedfile="simped.csv",
phen="CVD",outfile="simout.csv",covars=c("sex","age"),cov.int="age",
sep.ped=",",sep.phe=",",sep.gen=",")

## End(Not run)
```

---

geepack.lgst.int.batch.imputed

function to test gene-environment or gene-gene interactions between a dichotomous trait and a batch of imputed SNPs in families using Generalized Estimation Equation model

---

Description

Fit logistic regression via Generalized Estimation Equation (GEE) to test gene-environment or gene-gene interactions between a dichotomous phenotype and all imputed SNPs in a genotype file in family data under additive genetic model. The interaction term is the product of SNP genotype and a covariate for interaction (cov.int). The covariate for interaction (cov.int) can be SNP genotype (gene-gene interaction) or an environmental factor (gene-environment interaction). Only one interaction term is allowed. When cov.int is dichotomous, stratified analyses can be requested by specifying sub="Y". The covariance between the main effect (SNP) and the interaction effect is provided in the output when stratified analysis is not requested. Each family is treated as a cluster, with independence working correlation matrix used in the robust variance estimator. This function applies the same interaction test to all SNPs in the imputed genotype data. The interaction test is carried out by geepack.lgst.int.imputed function from GWAF where the the geese function from package geepack is used.

Usage

```
geepack.lgst.int.batch.imputed(genfile,phenfile,pedfile,outfile,phen,
covars,cov.int,sub="N",col.names=T,sep.ped=",",sep.phe=",",sep.gen=",")
```

Arguments

genfile	a character string naming the genotype file for reading (see format requirement in details)
phenfile	a character string naming the phenotype file for reading (see format requirement in details)
pedfile	a character string naming the pedigree file for reading (see format requirement in details)
outfile	a character string naming the result file for writing
phen	a character string for a phenotype name in phenfile
covars	a character vector for covariates in phenfile
cov.int	a character string naming the covariate for interaction, the covariate has to be included in covars
sub	"N" (default) for no stratified analysis, and "Y" for requesting stratified analyses (only when cov.int is dichotomous)
col.names	a logical value indicating whether the output file should contain column names
sep.ped	the field separator character for pedigree file
sep.phe	the field separator character for phenotype file
sep.gen	the field separator character for genotype file

Details

Similar to the details for geepack.lgst.int.batch but here the SNP data contains imputed genotypes (allele dosages) that are continuous and range from 0 to 2.

Value

No value is returned. Instead, results are written to outfile. If stratified analyses are requested, the result file will include the following columns. Otherwise, cov_beta_snp_beta_int will be included instead of the results from stratified analyses, that is, beta_snp_cov0, se_snp_cov0, pval_snp_cov0, beta_snp_cov1, se_snp_cov1, and pval_snp_cov1.

phen	phenotype name
snp	SNP name
covar_int	the covariate for interaction
n	sample size used in analysis
AF	allele frequency of the coded allele
nd	the number of individuals in affected sample
AFd	allele frequency of the coded allele in affected sample

model	genetic model used in analysis, additive model only
beta_snp	regression coefficient of SNP covariate
se_snp	standard error of beta_snp
pval_snp	p-value of testing beta_snp not equal to zero
beta_snp_cov0	regression coefficient of SNP covariate in stratified analysis using the subset where cov.int level is 0
se_snp_cov0	standard error of beta_snp_cov0
pval_snp_cov0	p-value of testing beta_snp_cov0 not equal to zero
beta_snp_cov1	regression coefficient of SNP covariate in stratified analysis using the subset where cov.int level is 1
se_snp_cov1	standard error of beta_snp_cov1
pval_snp_cov1	p-value of testing beta_snp_cov1 not equal to zero
beta_int	regression coefficient of the interaction term
se_int	standard error of beta_int
pval_int	p-value of testing beta_int not equal to zero
remark	warning or additional information for the analysis, 'not converged' indicates the GEE analysis did not converge; 'logistic reg' indicates GEE model is replaced by logistic regression; 'exp count<5' indicates any expected count is less than 5 in phenotype-genotype table; 'not converged and exp count<5', 'logistic reg & exp count<5' are noted similarly; 'collinearity' indicates collinearity exists between SNP and some covariates

Author(s)

Qiong Yang <qyang@bu.edu> and Ming-Huei Chen <mhchen@bu.edu>

Examples

```
## Not run:
geepack.lgst.int.batch.imputed(phenfile="simphen.csv",genfile="simgen.csv",
pedfile="simped.csv",phen="CVD",outfile="simout.csv",covars=c("sex","age"),cov.int="sex",
sub="Y",sep.ped=",",sep.phe=",",sep.gen=",")

## End(Not run)
```

---

geepack.lgst.int.imputed

function for testing gene-environment or gene-gene interaction between a dichotomous trait and an imputed SNP in family data using Generalized Estimation Equation model

---

Description

Fit logistic regression via Generalized Estimation Equation (GEE) to test gene-environment or gene-gene interaction between a dichotomous phenotype and one imputed SNP in a genotype file under additive genetic model. The interaction term is the product of SNP genotype and a covariate for interaction (`cov.int`). The covariate for interaction (`cov.int`) can be SNP genotype (gene-gene interaction) or an environmental factor (gene-environment interaction). Only one interaction term is allowed. When `cov.int` is dichotomous, stratified analyses can be requested by specifying `sub="Y"`. The covariance between the main effect (SNP) and the interaction effect is provided in the output when stratified analysis is not requested. Each family is treated as a cluster, with independence working correlation matrix used in the robust variance estimator. This function is called in `geepack.lgst.int.batch.imputed` function to apply interaction test to all imputed SNPs in a genotype file. The interaction test is carried out by the `geese` function from package `geepack`.

Usage

```
geepack.lgst.int.imputed(snp,phen,test.dat,covar,cov.int,sub="N")
```

Arguments

`snp`	genotype data of a SNP
`phen`	a character string for a phenotype name in `test.dat`
`test.dat`	the product of merging phenotype, genotype and pedigree data, should be ordered by "famid"
`covar`	a character vector for covariates in `test.dat`
`cov.int`	a character string naming the covariate for interaction, the covariate has to be included in `covar`
`sub`	"N" (default) for no stratified analysis, and "Y" for requesting stratified analyses (only when `cov.int` is dichotomous)

Details

Similar to the details for `geepack.lgst.int.batch` function but here the SNP data contains imputed genotypes (allele dosages) that are continuous and range from 0 to 2.

Value

Please see value in `geepack.lgst.int.batch.imputed` function.

Author(s)

Qiong Yang <qyang@bu.edu> and Ming-Huei Chen <mhchen@bu.edu>

References

Liang, K.Y. and Zeger, S.L. (1986) Longitudinal data analysis using generalized linear models. *Biometrika*, **73** 13–22.

Zeger, S.L. and Liang, K.Y. (1986) Longitudinal data analysis for discrete and continuous outcomes. *Biometrics*, **42** 121–130.

Yan, J and Fine, J. (2004) Estimating equations for association structures. *Stat Med*, **23** 859–874.

See Also

geese function from package geepack

Examples

```
## Not run:
geepack.lgst.int.imputed(snp=data[,"rs123"],phen="CVD",test.dat=data,covar=c("age",sex"),
cov.int="sex",sub="Y")

## End(Not run)
```

---

geepack.quant.batch *function to test genetic associations between a continuous trait and a batch of genotyped SNPs in families using Generalized Estimation Equation model*

Description

Fit Generalized Estimation Equation (GEE) model to test associations between a continuous phenotype and all genotyped SNPs in a genotype file in family data with user specified genetic model. Each pedigree is treated as a cluster, with independence working correlation matrix used in the robust variance estimator. The proportion of phenotype variation explained by the tested SNP is not provided. This function applies the same trait-SNP association test to all genotyped SNPs in the genotype data. The trait-SNP association test is carried out by using the geese function from package geepack.

Usage

```
geepack.quant.batch(phenfile,genfile,pedfile,phen,model="a",covars=NULL,outfile,
col.names=T,sep.ped=",",sep.phe=",",sep.gen=",")
```

Arguments

genfile	a character string naming the genotype file for reading (see format requirement in details)
phenfile	a character string naming the phenotype file for reading (see format requirement in details)
pedfile	a character string naming the pedigree file for reading (see format requirement in details)
outfile	a character string naming the result file for writing
phen	a character string for a phenotype name in phenfile
covars	a character vector for covariates in phenfile

model	a single character of 'a','d','g', or 'r', with 'a'=additive, 'd'=dominant, 'g'=general and 'r'=recessive models
col.names	a logical value indicating whether the output file should contain column names
sep.ped	the field separator character for pedigree file
sep.phe	the field separator character for phenotype file
sep.gen	the field separator character for genotype file

Details

For a continuous trait, the geepack.quant.batch function first reads in and merges phenotype-covariates, genotype and pedigree files, then tests the association of phen against all SNPs in genfile. genfile contains unique individual id and genotype data, with the column names being "id" and SNP names. For each SNP, the genotype data should be coded as 0, 1, 2 indicating the numbers of the coded alleles. The SNP name in genotype file should not have any dash, '-' and other special characters(dots and underscores are OK). phenfile contains unique individual id, phenotype and covariates data, with the column names being "id" and phenotype and covaraite names. pedfile contains pedigree informaion, with the column names being "famid","id","fa","mo","sex". In all files, missing value should be an empty space, except missing parental id in pedfile. SNPs with low genotype counts (especially minor allele homozygote) may be omitted or analyzed with dominant model. The geepack.quant.batch function fits GEE model using each pedigree as a cluster with geese function from geepack package.

Value

No value is returned. Instead, results are written to outfile. When the genetic model is 'a', 'd' or 'r', the result includes the following columns. When the genetic model is 'g', beta and se are replaced with beta10, beta20, beta21, se10, se20, se21 .

phen	phenotype name
snp	SNP name
n0	the number of individuals with 0 copy of coded alleles
n1	the number of individuals with 1 copy of coded alleles
n2	the number of individuals with 2 copies of coded alleles
beta	regression coefficient of SNP covariate
se	standard error of beta
chisq	Chi-square statistic for testing beta not equal to zero
df	degree of freedom of the Chi-square statistic
model	model actually used in the analysis
pval	p-value of the chi-square statistic
beta10	regression coefficient of genotype with 1 copy of coded allele vs. that with 0 copy
beta20	regression coefficient of genotype with 2 copy of coded allele vs. that with 0 copy

beta21	regression coefficient of genotype with 2 copy of coded allele vs. that with 1 copy
se10	standard error of beta10
se20	standard error of beta20
se21	standard error of beta21

Author(s)

Qiong Yang <qyang@bu.edu> and Ming-Huei Chen <mhchen@bu.edu>

References

Liang, K.Y. and Zeger, S.L. (1986) Longitudinal data analysis using generalized linear models. *Biometrika*, **73** 13–22.

Zeger, S.L. and Liang, K.Y. (1986) Longitudinal data analysis for discrete and continuous outcomes. *Biometrics*, **42** 121–130.

Yan, J and Fine, J. (2004) Estimating equations for association structures. *Stat Med*, **23** 859–874.

Examples

```
## Not run:
geepack.quant.batch(phenfile="simphen.csv",genfile="simgen.csv",pedfile="simped.csv",
phen="SIMQT",model="a",outfile="simout.csv",sep.ped=",",sep.phe=",",sep.gen=",")

## End(Not run)
```

---

geepack.quant.batch.imputed

function to test associations between a continuous trait and a batch of imputed SNPs in families using Generalized Estimation Equation model

---

Description

Fit Generalized Estimation Equation (GEE) model to test associations between a continuous phenotype and all imputed SNPs in a genotype file in family data under additive genetic model. Each family is treated as a cluster, with independence working correlation matrix used in the robust variance estimator. The proportion of phenotype variation explained by the tested SNP is not provided. This function applies the same trait-SNP association test to all imputed SNPs in the genotype data. The trait-SNP association test is carried out by using the geese function from package geepack.

Usage

```
geepack.quant.batch.imputed(phenfile,genfile,pedfile,phen,
covars=NULL,outfile,col.names=T,sep.ped=",",sep.phe=",",sep.gen=",")
```

Arguments

phenfile	a character string naming the phenotype file for reading (see format requirement in details)
genfile	a character string naming the genotype file for reading (see format requirement in details)
pedfile	a character string naming the pedigree file for reading (see format requirement in details)
phen	a character string for a phenotype name in phenfile
covars	a character vector for covariates in phenfile
outfile	a character string naming the result file for writing
col.names	a logical value indicating whether the output file should contain column names
sep.ped	the field separator character for pedigree file
sep.phe	the field separator character for phenotype file
sep.gen	the field separator character for genotype file

Details

Similar to the details for geepack.quant.batch function but here the SNP data contains imputed genotypes (allele dosages) that are continuous and range from 0 to 2. In addition, the user specified genetic model argument is not available.

Value

No value is returned. Instead, results are written to outfile.

phen	phenotype name
snp	SNP name
N	the number of individuals in analysis
AF	imputed allele frequency of coded allele
beta	regression coefficient of SNP covariate
se	standard error of beta
pval	p-value of testing beta not equal to zero

Author(s)

Qiong Yang <qyang@bu.edu> and Ming-Huei Chen <mhchen@bu.edu>

References

Liang, K.Y. and Zeger, S.L. (1986) Longitudinal data analysis using generalized linear models. *Biometrika*, **73** 13–22.

Zeger, S.L. and Liang, K.Y. (1986) Longitudinal data analysis for discrete and continuous outcomes. *Biometrics*, **42** 121–130.

Yan, J and Fine, J. (2004) Estimating equations for association structures. *Stat Med*, **23** 859–874.

Examples

```
## Not run:
geepack.quant.batch.imputed(phenfile="simphen.csv",genfile="simgen.csv",
pedfile="simped.csv",phen="SIMQT",outfile="simout.csv",col.names=T,covars="sex",
sep.ped=",",sep.phe=",",sep.gen=",")

## End(Not run)
```

`geepack.quant.int.batch`

function to test gene-environment or gene-gene interactions for a continuous trait and a batch of genotyped SNPs in families using Generalized Estimation Equation model

---

Description

Fit Generalized Estimation Equation (GEE) model to test gene-environment or gene-gene interactions for a continuous phenotype and all genotyped SNPs in a genotype file in family data under additive genetic model. The interaction term is the product of SNP genotype and a covariate for interaction (`cov.int`). The covariate for interaction (`cov.int`) can be SNP genotype (gene-gene interaction) or an environmental factor (gene-environment interaction). Only one interaction term is allowed. When `cov.int` is dichotomous, stratified analyses can be requested by specifying `sub="Y"`. The covariance between the main effect (SNP) and the interaction effect is provided in the output when stratified analysis is not requested. Each pedigree is treated as a cluster, with independence working correlation matrix used in the robust variance estimator. This function applies the same interaction test to all genotyped SNPs in the genotype data. In each test for interaction, the `geese` function from `geepack` package is used.

Usage

```
geepack.quant.int.batch(phenfile,genfile,pedfile,phen,covars,cov.int,sub="N",outfile,
col.names=T,sep.ped=",",sep.phe=",",sep.gen=",")
```

Arguments

`genfile`	a character string naming the genotype file for reading (see format requirement in details)
`phenfile`	a character string naming the phenotype file for reading (see format requirement in details)
`pedfile`	a character string naming the pedigree file for reading (see format requirement in details)
`outfile`	a character string naming the result file for writing
`phen`	a character string for a phenotype name in `phenfile`
`covars`	a character vector for covariates in `phenfile`
`cov.int`	a character string naming the covariate for interaction, the covariate has to be included in `covars`

sub	"N" (default) for no stratified analysis, and "Y" for requesting stratified analyses (only when cov.int is dichotomous)
col.names	a logical value indicating whether the output file should contain column names
sep.ped	the field separator character for pedigree file
sep.phe	the field separator character for phenotype file
sep.gen	the field separator character for genotype file

Details

For a continuous trait, the geepack.quant.int.batch function first reads in and merges phenotype-covariates, genotype and pedigree files, then tests gene-environment or gene-gene interaction and the association of phen against all genotyped SNPs in genfile. Only one interaction term is allowed, so is the covariate for interaction (cov.int). When cov.int is dichotomous, stratified analyses can be requested by specifying sub="Y". The covariance between the main effect (SNP) and the interaction effect is provided in the output when stratified analysis is not requested. genfile contains unique individual id and genotype data, with the column names being "id" and SNP names. For each SNP, the genotype data should be coded as 0, 1, 2 indicating the numbers of the coded alleles. The SNP name in genotype file should not have any dash, '-' and other special characters(dots and underscores are OK). phenfile contains unique individual id, phenotype and covariate data, with the column names being "id" and phenotype and covaraite names. pedfile contains pedigree informaion, with the column names being "famid","id","fa","mo","sex". In all files, missing value should be an empty space, except missing parental id in pedfile. SNPs with low genotype counts (especially minor allele homozygote) may be omitted. The geepack.quant.int.batch function fits GEE model using geese function from geepack package.

Value

No value is returned. Instead, results are written to outfile. If stratified analyses are requested, the result file will include the following columns. Otherwise, cov_beta_snp_beta_int will be included instead of the results from stratified analyses, that is, beta_snp_cov0, se_snp_cov0, pval_snp_cov0, beta_snp_cov1, se_snp_cov1, and pval_snp_cov1.

phen	phenotype name
snp	SNP name
covar_int	the covariate for interaction
n	sample size used in analysis
AF	allele frequency of the coded allele
model	genetic model used in analysis, additive model only
beta_snp	regression coefficient of SNP covariate
se_snp	standard error of beta_beta
pval_snp	p-value of testing beta_beta not equal to zero
beta_snp_cov0	regression coefficient of SNP covariate in stratified analysis using the subset where cov.int level is 0
se_snp_cov0	standard error of beta_snp_cov0
pval_snp_cov0	p-value of testing beta_snp_cov0 not equal to zero

beta_snp_cov1	regression coefficient of SNP covariate in stratified analysis using the subset where cov.int level is 1
se_snp_cov1	standard error of beta_snp_cov1
pval_snp_cov1	p-value of testing beta_snp_cov1 not equal to zero
beta_int	regression coefficient of the interaction term
se_int	standard error of beta_int
pval_int	p-value of testing beta_int not equal to zero

Author(s)

Qiong Yang <qyang@bu.edu> and Ming-Huei Chen <mhchen@bu.edu>

References

Liang, K.Y. and Zeger, S.L. (1986) Longitudinal data analysis using generalized linear models. *Biometrika*, **73** 13–22.

Zeger, S.L. and Liang, K.Y. (1986) Longitudinal data analysis for discrete and continuous outcomes. *Biometrics*, **42** 121–130.

Yan, J and Fine, J. (2004) Estimating equations for association structures. *Stat Med*, **23** 859–874.

Examples

```
## Not run:
geepack.quant.int.batch(phenfile="simphen.csv",genfile="simgen.csv",
pedfile="simped.csv",phen="SIMQT",outfile="simout.csv",col.names=T,covars="age",
cov.int="age",sep.ped=",",sep.phe=",",sep.gen=",")

## End(Not run)
```

---

geepack.quant.int.batch.imputed

function to test gene-environment or gene-gene interactions between a continuous trait and a batch of imputed SNPs in families using Generalized Estimation Equation model

---

Description

Fit Generalized Estimation Equation (GEE) model to test gene-environment or gene-gene interactions for a continuous phenotype and all imputed SNPs in a genotype file in family data under additive genetic model. The interaction term is the product of SNP genotype (allelic dosage) and a covariate for interaction (cov.int). The covariate for interaction (cov.int) can be SNP genotype (gene-gene interaction) or an environmental factor (gene-environment interaction). Only one interaction term is allowed. When cov.int is dichotomous, stratified analyses can be requested by specifying sub="Y". The covariance between the main effect (SNP) and the interaction effect is provided in the output when stratified analysis is not requested. Each pedigree is treated as a cluster, with independence working correlation matrix used in the robust variance estimator. This function applies the same interaction test to all imputed SNPs in the genotype data. In each test for interaction, the geese function from geepack package is used.

Usage

```
geepack.quant.int.batch.imputed(phenfile,genfile,pedfile,phen,covars,cov.int,sub="N",
outfile,col.names=T,sep.ped=",",sep.phe=",",sep.gen=",")
```

Arguments

phenfile	a character string naming the phenotype file for reading (see format requirement in details)
genfile	a character string naming the (imputed) genotype file for reading (see format requirement in details)
pedfile	a character string naming the pedigree file for reading (see format requirement in details)
outfile	a character string naming the result file for writing
phen	a character string for a phenotype name in phenfile
covars	a character vector for covariates in phenfile
cov.int	a character string naming the covariate for interaction, the covariate has to be included in covars
sub	"N" (default) for no stratified analysis, and "Y" for requesting stratified analyses (only when cov.int is dichotomous)
col.names	a logical value indicating whether the output file should contain column names
sep.ped	the field separator character for pedigree file
sep.phe	the field separator character for phenotype file
sep.gen	the field separator character for genotype file

Details

Similar to the details for geepack.quant.int.batch function but here the SNP data contains imputed genotypes (allele dosages) that are continuous and range from 0 to 2.

Value

Please see value in geepack.quant.int.batch function.

Author(s)

Qiong Yang <qyang@bu.edu> and Ming-Huei Chen <mhchen@bu.edu>

References

Liang, K.Y. and Zeger, S.L. (1986) Longitudinal data analysis using generalized linear models. *Biometrika*, **73** 13–22.

Zeger, S.L. and Liang, K.Y. (1986) Longitudinal data analysis for discrete and continuous outcomes. *Biometrics*, **42** 121–130.

Yan, J and Fine, J. (2004) Estimating equations for association structures. *Stat Med*, **23** 859–874.

Examples

```
## Not run:
geepack.quant.int.batch.imputed(phenfile="simphen.csv",genfile="simgen.csv",
pedfile="simped.csv",phen="SIMQT",outfile="simout.csv",col.names=T,covars=c("sex",age"),
cov.int="sex",sub="Y",sep.ped=",",sep.phe=",",sep.gen=",")

## End(Not run)
```

`glmm.lgst` *function for testing association between a dichotomous trait and a genotyped SNP in family data using Generalized Linear Mixed Effects model*

Description

Fit a Generalized Linear Mixed Effects model (GLMM) with logistic link and a normal distributed random intercept for each cluster to test association between a dichotomous phenotype and one genotyped SNP in a genotype file with user specified genetic model. Each family is treated as a cluster. When analyze rare variants for dichotomous traits, GLMM, as implemeted by this function, is recommended over other methods such as GEE. The trait-SNP association test is carried out by the `lmer` function from package `lme4`. This function is called in `glmm.lgst.batch` function to apply association test to all SNPs in the genotype data.

Usage

```
glmm.lgst(snp, phen, test.dat, covar = NULL, model = "a")
```

Arguments

`snp`	genotype data of a SNP
`phen`	a character string for a phenotype name in `test.dat`
`test.dat`	the product of merging phenotype, genotype and pedigree data, should be ordered by "famid"
`covar`	a character vector for covariates in `test.dat`
`model`	a single character of 'a','d','g', or 'r', with 'a'=additive, 'd'=dominant, 'g'=general and 'r'=recessive models

Details

The `glmm.lgst` function tests association between a dichtomous trait and a SNP from a dataset that contains phenotype, genotype and pedigree data (test.dat), where the dataset needs to be ordered by famid.

Value

Please see output in `glmm.lgst.batch`.

Author(s)

Qiong Yang <qyang@bu.edu> and Ming-Huei Chen <mhchen@bu.edu>

References

Bates D, Maechler M and Bolker B. (2010) lme4: Linear mixed-effects models using S4 classes Longitudinal data analysis using generalized linear models. http://cran.r-project.org/web/packages/lme4/.

Chen MH, Liu X, Wei F, Larson MG, Fox CS, Vasan RS and Yang Q. (2011) A comparison of strategies for analyzing dichotomous outcomes in genome-wide association studies with general pedigrees. Genetic Epidemiology 35:650-657.

See Also

`lmer` function from package `lme4`

Examples

```
## Not run:
glmm.lgst(snp=data[,"rs123"],phen="CVD",test.dat=data,model="a",covar=c("age","sex"))

## End(Not run)
```

---

glmm.lgst.batch *function to test genetic association between a dichotomous trait and a batch of genotyped SNPs in families using Generalized Linear Mixed Effects model*

Description

Fit Generalized Linear Mixed Effects model (GLMM) with logistic link and a normal distributed random intercept for each cluster to test associations between a dichotomous phenotype and all genotyped SNPs in a genotype file in family data with user specified genetic model. Each pedigree is treated as a cluster. This function applies the same trait-SNP association test to all SNPs in the genotype data. When analyzing rare variants for dichotomous traits, this GLMM, as implemeted by this function, is recommended over other methods such as GEE. The trait-SNP association test is carried out by `glmm.lgst` function where the the `lmer` function from package `lme4` is used.

Usage

```
glmm.lgst.batch(genfile, phenfile, pedfile, outfile, phen, covars = NULL,
model = "a", col.names = T, sep.ped = ",", sep.phe = ",", sep.gen = ",")
```

Arguments

genfile	a character string naming the genotype file for reading (see format requirement in details)
phenfile	a character string naming the phenotype file for reading (see format requirement in details)
pedfile	a character string naming the pedigree file for reading (see format requirement in details)
outfile	a character string naming the result file for writing
phen	a character string for a phenotype name in phenfile
covars	a character vector for covariates in phenfile
model	a single character of 'a','d','g', or 'r', with 'a'=additive, 'd'=dominant, 'g'=general and 'r'=recessive models
col.names	a logical value indicating whether the output file should contain column names
sep.ped	the field separator character for pedigree file
sep.phe	the field separator character for phenotype file
sep.gen	the field separator character for genotype file

Details

The glmm.lgst.batch function first reads in and merges phenotype-covariates, genotype and pedigree files, then tests the association of phen against all SNPs in genfile. genfile contains unique individual id and genotype data, with the column names being "id" and SNP names. For each genotyped SNP, the genotype data should be coded as 0, 1, 2 indicating the numbers of the coded alleles. The SNP names in genotype file should not have any dash, '-' and other special characters(dots and underscores are OK). phenfile contains unique individual id, phenotype and covariates data, with the column names being "id" and phenotype and covaraite names. pedfile contains pedigree informaion, with the column names being "famid","id","fa","mo","sex". In all files, missing value should be an empty space, except missing parental id in pedfile. Only phenotypes with two categories are analyzed. A phenotype should be coded as 0 and 1, with 1 denoting affected and 0 unaffected. SNPs with low genotype counts (especially minor allele homozygote) may be omitted or analyzed with dominant model or analyzed with logistic regression. The glmm.lgst.batch function fits GLMM using each pedigree as a cluster with glmm.lgst function from GWAF package and lmer function from lme4 package.

Value

No value is returned. Instead, results are written to outfile. When the genetic model is 'a', 'd' or 'r', the result includes the following columns. When the genetic model is 'g', beta and se are replaced with beta10, beta20, beta21, se10, se20, and se21 .

phen	phenotype name
snp	SNP name
n0	the number of individuals with 0 copy of coded alleles
n1	the number of individuals with 1 copy of coded alleles

n2	the number of individuals with 2 copies of coded alleles
nd0	the number of individuals with 0 copy of coded alleles in affected sample
nd1	the number of individuals with 1 copy of coded alleles in affected sample
nd2	the number of individuals with 2 copies of coded alleles in affected sample
miss.0	Genotype missing rate in unaffected sample
miss.1	Genotype missing rate in affected sample
miss.diff.p	P-value of differential missingness test between unaffected and affected samples
beta	regression coefficient of SNP covariate
se	standard error of beta
chisq	Chi-square statistic for testing beta not equal to zero
df	degree of freedom of the Chi-square statistic
model	model actually used in the analysis
remark	warning or additional information for the analysis, 'exp count<5' indicates any expected count is less than 5 in phenotype-genotype table; 'collinearity' indicates collinearity exists between SNP and some covariates
pval	p-value of the chi-square statistic
beta10	regression coefficient of genotype with 1 copy of coded allele vs. that with 0 copy
beta20	regression coefficient of genotype with 2 copy of coded allele vs. that with 0 copy
beta21	regression coefficient of genotype with 2 copy of coded allele vs. that with 1 copy
se10	standard error of beta10
se20	standard error of beta20
se21	standard error of beta21

Author(s)

Qiong Yang <qyang@bu.edu> and Ming-Huei Chen <mhchen@bu.edu>

Examples

```
## Not run:
glmm.lgst.batch(phenfile="simphen.csv",genfile="simgen.csv",pedfile="simped.csv",
phen="SIMQT",model="d",outfile="simout.csv",sep.ped=",",sep.phe=",",sep.gen=",")

## End(Not run)
```

GWplot *function for making genome-wide p-value plot*

Description

GWplot function plots -log_10 p-value based on SNP's chromosomal position in bitmap format.

Usage

```
GWplot(data, pval, pos, chr, chr.plot = c(1:22, "X"), title.text = "",
ylim = Inf, outfile, cutoff1 = 5e-08, cutoff2 = 4e-07)
```

Arguments

data	a dataframe that contains p-values, chromosome number and physical position of SNPs
pval	a character string correspong to the name of the p-value column
pos	a character string correspong to the name of column with SNP physical positions
chr	a character string correspong to the name of column with SNP chromosome number
chr.plot	the chsomosomes of interest for GWplot; either 1:22 or c(1:22,"X"), default chr.plot=c(1:22,"X"), "X" for X chromosome
title.text	the title of the genome-wide p-value plot
ylim	the maximum of -log_10 p-value to be plotted, useful when not want to plot extremely small p-values
outfile	the file name (xxxx.bmp) for output plot in bitmap format
cutoff1	genome-wide significance; default is 5E-8 ; p-values below this threshold will be highlighted in red
cutoff2	suggestive genome-wide significance; default is 4E-7; p-values below this threshold but above cutoff1 will be highlighted in blue

Details

When the dataset has 0 p-value, GWplot will generate pvalzero.csv that contain the results with 0 p-value and make the genome-wide p-value plot by replacing 0 p-value with 5E-324. P-values that reach genome-wide significance are displayed in red color; P-values that reach suggestive genome-wide significance but not genome-wide significance are displayed in blue color.

Author(s)

Qiong Yang <qyang@bu.edu> and Ming-Huei Chen <mhchen@bu.edu>

Examples

```
## Not run:
GWplot(data=result,pval="pval",pos="position",chr="chr",chr.plot=c(1:22,"X"),outfile="GWP_test.bmp")

## End(Not run)
```

`lmepack.batch` *function to test genetic associations between a continuous trait and a batch of genotyped SNPs in families using Linear Mixed Effects model*

Description

Fit linear mixed effects (LME) model to test associations between a continuous phenotype and all SNPs in a genotype file in family data under user specified genetic model. The SNP genotype is treated as a fixed effect, and a random effect correlated according to degree of relatedness within a family is also fitted. In each trait-SNP assocaition test, the `lmekin` function from package `coxme` is used.

Usage

```
lmepack.batch(phenfile, genfile, pedfile, phen, kinmat, model = "a", covars = NULL,
outfile, col.names = T, sep.ped = ",", sep.phe = ",", sep.gen = ",")
```

Arguments

`genfile`	a character string naming the genotype file for reading (see format requirement in details)
`phenfile`	a character string naming the phenotype file for reading (see format requirement in details)
`pedfile`	a character string naming the pedigree file for reading (see format requirement in details)
`outfile`	a character string naming the result file for writing
`phen`	a character string for a phenotype name in `phenfile`
`covars`	a character vector for covariates in `phenfile`
`model`	a single character of 'a','d','g', or 'r', with 'a'=additive, 'd'=dominant, 'g'=general and 'r'=recessive models
`kinmat`	a character string naming the file where kinship coefficient matrix is kept
`col.names`	a logical value indicating whether the output file should contain column names
`sep.ped`	the field separator character for pedigree file
`sep.phe`	the field separator character for phenotype file
`sep.gen`	the field separator character for genotype file

Details

The lmepack.batch function first reads in and merges phenotype-covariates, genotype and pedigree files, then tests the association of phen against all SNPs in genfile. genfile contains unique individual id and genotype data, with the column names being "id" and SNP names. For each SNP, the genotype data should be coded as 0, 1, 2 indicating the numbers of the coded alleles. The SNP name in genotype file should not have any dash, '-' and other special characters(dots and underscores are OK). phenfile contains unique individual id, phenotype and covariates data, with the column names being "id" and phenotype and covaraite names. pedfile contains pedigree informaion, with the column names being "famid","id","fa","mo","sex". In all files, missing value should be an empty space, except missing parental id in pedfile. SNPs with low genotype counts (especially minor allele homozygote) may be omitted or analyzed with dominant model. The lmepack.batch function fits LME model using a modified lmekin function from coxme package.

Value

No value is returned. Instead, results are written to outfile. When the genetic model is 'a', 'd' or 'r', the result includes the following columns. When the genetic model is 'g', beta and se are replaced with beta10, beta20, beta21, se10, se20, se21.

phen	phenotype name
snp	SNP name
n0	the number of individuals with 0 copy of coded alleles
n1	the number of individuals with 1 copy of coded alleles
n2	the number of individuals with 2 copies of coded alleles
h2q	the portion of phenotypic variation explained by the SNP
beta	regression coefficient of SNP covariate
se	standard error of beta
chisq	Chi-square statistic for testing beta not equal to zero
df	degree of freedom of the Chi-square statistic
model	model actually used in the analysis
pval	p-value of the chi-square statistic
beta10	regression coefficient of genotype with 1 copy of coded allele vs. that with 0 copy
beta20	regression coefficient of genotype with 2 copy of coded allele vs. that with 0 copy
beta21	regression coefficient of genotype with 2 copy of coded allele vs. that with 1 copy
se10	standard error of beta10
se20	standard error of beta20
se21	standard error of beta21

Author(s)

Qiong Yang <qyang@bu.edu> and Ming-Huei Chen <mhchen@bu.edu>

References

coxme package: mixed-effects Cox models, sparse matrices, and modeling data from large pedigrees. Beth Atkinson (atkinson@mayo.edu) for pedigree functions.Terry Therneau (therneau@mayo.edu) for all other functions. 2007. Ref Type: Computer Programhttp://cran.r-project.org/web/packages/coxme/.

Abecasis, G. R., Cardon, L. R., Cookson, W. O., Sham, P. C., & Cherny, S. S. Association analysis in a variance components framework. *Genet Epidemiol*, **21** Suppl 1, S341-S346 (2001).

Examples

```
## Not run:
lmepack.batch(phenfile="simphen.csv",genfile="simgen.csv",pedfile="simped.csv",
phen="SIMQT",kinmat="simkmat.Rdata",model="a",outfile="simout.csv",col.names=T,
sep.ped=",",sep.phe=",",sep.gen=",")

## End(Not run)
```

---

lmepack.batch.imputed *function to test associations between a continuous trait and a batch of imputed SNPs in families using Linear Mixed Effects model*

Description

Fit linear mixed effects (LME) model to test associations between a continuous phenotype and all imputed SNPs in a genotype file in family data under additive genetic model. The SNP genotype is treated as a fixed effect, and a random effect correlated according to degree of relatedness within a family is also fitted. In each trait-SNP assocaition test, the `lmekin` function from package `coxme` is used.

Usage

```
lmepack.batch.imputed(phenfile, genfile, pedfile, phen, kinmat, covars = NULL,
outfile, col.names = T, sep.ped = ",", sep.phe = ",", sep.gen = ",")
```

Arguments

`phenfile`	a character string naming the phenotype file for reading (see format requirement in details)
`genfile`	a character string naming the genotype file for reading (see format requirement in details)
`pedfile`	a character string naming the pedigree file for reading (see format requirement in details)
`phen`	a character string for a phenotype name in `phenfile`

kinmat	a character string naming the file where kinship coefficient matrix is kept
covars	a character vector for covariates in phenfile
outfile	a character string naming the result file for writing
col.names	a logical value indicating whether the output file should contain column names
sep.ped	the field separator character for pedigree file
sep.phe	the field separator character for phenotype file
sep.gen	the field separator character for genotype file

Details

Similar to the details for lmepack.batch function but here the SNP data contains imputed genotypes (allele dosages) that are continuous and range from 0 to 2. In addition, the user specified genetic model argument is not available.

Value

No value is returned. Instead, results are written to outfile.

phen	phenotype name
snp	SNP name
N	the number of individuals in analysis
AF	imputed allele frequency of coded allele
h2q	the portion of phenotypic variation explained by the SNP
beta	regression coefficient of SNP covariate
se	standard error of beta
pval	p-value of testing beta not equal to zero

Author(s)

Qiong Yang <qyang@bu.edu> and Ming-Huei Chen <mhchen@bu.edu>

References

coxme package: mixed-effects Cox models, sparse matrices, and modeling data from large pedigrees. Beth Atkinson (atkinson@mayo.edu) for pedigree functions.Terry Therneau (therneau@mayo.edu) for all other functions. 2007. Ref Type: Computer Program http://cran.r-project.org/web/packages/coxme/.

Abecasis, G. R., Cardon, L. R., Cookson, W. O., Sham, P. C., & Cherny, S. S. Association analysis in a variance components framework. *Genet Epidemiol*, **21** Suppl 1, S341-S346 (2001).

Examples

```
## Not run:
lmepack.batch.imputed(phenfile="simphen.csv",genfile="simgen.csv",pedfile="simped.csv",
phen="SIMQT",kinmat="simkmat.Rdata",outfile="simout.csv",covars=c("age","sex"),
sep.ped=",",sep.phe=",",sep.gen=",")

## End(Not run)
```

`lmepack.int.batch` *function to test gene-environment or gene-gene interactions for a continuous trait and a batch of genotyped SNPs in families using Linear Mixed Effects model*

Description

Fit linear mixed effects model (LME) to test gene-environment or gene-gene interactions for a continuous phenotype and all SNPs in a genotype file in family data under additive genetic model. The interaction term is the product of SNP genotype and a covariate for interaction (`cov.int`). The covariate for interaction (`cov.int`) can be SNP genotype (gene-gene interaction) or an environmental factor (gene-environment interaction). Only one interaction term is allowed. When `cov.int` is dichotomous, stratified analyses can be requested by specifying `sub="Y"`. The covariance between the main effect (SNP) and the interaction effect is provided in the output when stratified analysis is not requested. The SNP genotype and the interaction are treated as fixed effects, and a random effect correlated according to degree of relatedness within a family is also fitted. In each test for interaction, the `lmekin` function from package `coxme` is used.

Usage

```
lmepack.int.batch(phenfile,genfile,pedfile,phen,kinmat,covars,cov.int,sub="N",
outfile,col.names=T,sep.ped=",",sep.phe=",",sep.gen=",")
```

Arguments

`genfile`	a character string naming the genotype file for reading (see format requirement in details)
`phenfile`	a character string naming the phenotype file for reading (see format requirement in details)
`pedfile`	a character string naming the pedigree file for reading (see format requirement in details)
`outfile`	a character string naming the result file for writing
`phen`	a character string for a phenotype name in `phenfile`
`covars`	a character vector for covariates in `phenfile`
`cov.int`	a character string naming the covariate for interaction, the covariate has to be included in `covars`
`sub`	"N" (default) for no stratified analysis, and "Y" for requesting stratified analyses (only when `cov.int` is dichotomous)
`kinmat`	a character string naming the file where kinship coefficient matrix is kept
`col.names`	a logical value indicating whether the output file should contain column names
`sep.ped`	the field separator character for pedigree file
`sep.phe`	the field separator character for phenotype file
`sep.gen`	the field separator character for genotype file

Details

The lmepack.int.batch function first reads in and merges phenotype-covariates, genotype and pedigree files, then tests gene-environment or gene-gene interaction for phen against all SNPs in genfile. Only one interaction term is allowed, so is the covariate for interaction (cov.int). When cov.int is dichotomous, stratified analyses can be requested by specifying sub="Y". The covariance between the main effect (SNP) and the interaction effect is provided in the output when stratified analysis is not requested. genfile contains unique individual id and genotype data, with the column names being "id" and SNP names. For each SNP, the genotype data should be coded as 0, 1, 2 indicating the numbers of the coded alleles. The SNP name in genotype file should not have any dash, '-' and other special characters(dots and underscores are OK). phenfile contains unique individual id, phenotype and covariate data, with the column names being "id" and phenotype and covaraite names. pedfile contains pedigree informaion, with the column names being "famid","id","fa","mo","sex". In all files, missing value should be an empty space, except missing parental id in pedfile. SNPs with low genotype counts (especially minor allele homozygote) may be omitted. The lmepack.int.batch function fits LME model using lmekin function from coxme package.

Value

No value is returned. Instead, results are written to outfile. If stratified analyses are requested, the result file will include the following columns. Otherwise, cov_beta_snp_beta_int will be included instead of the results from stratified analyses, that is, (beta_snp_cov0, se_snp_cov0, pval_snp_cov0, beta_snp_cov1, se_snp_cov1, and pval_snp_cov1).

phen	phenotype name
snp	SNP name
covar_int	the covariate for interaction
n	sample size used in analysis
AF	allele frequency of the coded allele
model	genetic model used in analysis, additive model only
beta_snp	regression coefficient of SNP covariate
se_snp	standard error of beta_snp
pval_snp	p-value of testing beta_snp not equal to zero
beta_snp_cov0	regression coefficient of SNP covariate in stratified analysis using the subset where cov.int level is 0
se_snp_cov0	standard error of beta_snp_cov0
pval_snp_cov0	p-value of testing beta_snp_cov0 not equal to zero
beta_snp_cov1	regression coefficient of SNP covariate in stratified analysis using the subset where cov.int level is 1
se_snp_cov1	standard error of beta_snp_cov1
pval_snp_cov1	p-value of testing beta_snp_cov1 not equal to zero
beta_int	regression coefficient of the interaction term
se_int	standard error of beta_int
pval_int	p-value of testing beta_int not equal to zero

Author(s)

Qiong Yang <qyang@bu.edu> and Ming-Huei Chen <mhchen@bu.edu>

References

coxme package: mixed-effects Cox models, sparse matrices, and modeling data from large pedigrees. Beth Atkinson (atkinson@mayo.edu) for pedigree functions.Terry Therneau (therneau@mayo.edu) for all other functions. 2007. Ref Type: Computer Program http://cran.r-project.org/.

Abecasis, G. R., Cardon, L. R., Cookson, W. O., Sham, P. C., & Cherny, S. S. Association analysis in a variance components framework. *Genet Epidemiol*, **21** Suppl 1, S341-S346 (2001).

Examples

```
## Not run:
lmepack.int.batch(phenfile="simphen.csv",genfile="simgen.csv",pedfile="simped.csv",
phen="SIMQT",kinmat="simkmat.Rdata",outfile="simout.csv",covars=c("age","sex"),
cov.int="sex",sub="Y",sep.ped=",",sep.phe=",",sep.gen=",")

## End(Not run)
```

---

lmepack.int.batch.imputed

function to test gene-environment or gene-gen interaction and associations between a continuous trait and a batch of imputed SNPs in families using Linear Mixed Effects model

---

Description

Fit linear mixed effect model to test gene-environment or gene-gen interactions and genetic associations for a continuous phenotype and all imputed SNPs in a genotype file under additive genetic model. The interaction term is the product of SNP genotype (allelic dosage) and a covariate for interaction (`cov.int`). The covariate for interaction (`cov.int`) can be SNP genotype (gene-gene interaction) or an environmental factor (gene-environment interaction). Only one interaction term is allowed. When (`cov.int`) is dichotomous, stratified analyses can be requested by specifying `sub="Y"`. The covariance between the main effect (SNP) and the interaction effect is provided in the output when stratified analysis is not requested. The SNP genotype and the interaction are treated as fixed effectd, and a random effect correlated according to degree of relatedness within a family is also fitted. In each test for trait-SNP assocaition or interaction, the `lmekin()` function from package `coxme` is used.

Usage

```
lmepack.int.batch.imputed(phenfile,genfile,pedfile,phen,kinmat,covars,cov.int,sub="N",
outfile,col.names=T,sep.ped=",",sep.phe=",",sep.gen=",")
```

Arguments

phenfile	a character string naming the phenotype file for reading (see format requirement in details)
genfile	a character string naming the (imputed) genotype file for reading (see format requirement in details)
pedfile	a character string naming the pedigree file for reading (see format requirement in details)
outfile	a character string naming the result file for writing
phen	a character string for a phenotype name in phenfile
covars	a character vector for covariates in phenfile
cov.int	a character string naming the covariate for interaction, the covariate has to be included in covars
sub	"N" (default) for no stratified analysis, and "Y" for requesting stratified analyses (only when cov.int is dichotomous)
kinmat	a character string naming the file where kinship coefficient matrix is kept
col.names	a logical value indicating whether the output file should contain column names
sep.ped	the field separator character for pedigree file
sep.phe	the field separator character for phenotype file
sep.gen	the field separator character for genotype file

Details

Similar to the details for 'lmepack.int.batch' function but here the SNP data contains imputed genotypes (allele dosages) that are continuous and range from 0 to 2.

Value

Please see value in 'lmepack.int.batch' function.

Author(s)

Qiong Yang <qyang@bu.edu> and Ming-Huei Chen <mhchen@bu.edu>

References

coxme package: mixed-effects Cox models, sparse matrices, and modeling data from large pedigrees. Beth Atkinson (atkinson@mayo.edu) for pedigree functions.Terry Therneau (therneau@mayo.edu) for all other functions. 2007. Ref Type: Computer Program http://cran.r-project.org/.

Abecasis, G. R., Cardon, L. R., Cookson, W. O., Sham, P. C., & Cherny, S. S. Association analysis in a variance components framework. *Genet Epidemiol*, **21** Suppl 1, S341-S346 (2001).

Examples

```
## Not run:
lmepack.int.batch.imputed(phenfile="simphen.csv",genfile="simgen.csv",pedfile="simped.csv",
phen="SIMQT",kinmat="simkmat.Rdata",outfile="simout.csv",covars=c("age","sex"),
cov.int="sex",sub="Y",sep.ped=",",sep.phe=",",sep.gen=",")

## End(Not run)
```

---

lmeVpack.batch.imputed

function to efficiently test associations between a continuous trait and a batch of imputed SNPs in families using Linear Mixed Effects model

---

Description

A faster version of function `lmepack.batch.imputed`. Unlike `lmepack.batch.imputed` that estimates polygenic variation for every SNP in a batch of imputed SNPs, `lmeVpack.batch.imputed` only estimates once for a batch of imputed SNPs. Particularly recommended for analyzing 1000G imputed genotype data.

Usage

```
lmeVpack.batch.imputed(phenfile, genfile, pedfile, phen, kinmat, covars = NULL,
outfile, col.names = T, sep.ped = ",", sep.phe = ",", sep.gen = ",")
```

Arguments

`phenfile`	a character string naming the phenotype file for reading (see format requirement in details)
`genfile`	a character string naming the genotype file for reading (see format requirement in details)
`pedfile`	a character string naming the pedigree file for reading (see format requirement in details)
`phen`	a character string for a phenotype name in `phenfile`
`kinmat`	a character string naming the file where kinship coefficient matrix is kept
`covars`	a character vector for covariates in `phenfile`
`outfile`	a character string naming the result file for writing
`col.names`	a logical value indicating whether the output file should contain column names
`sep.ped`	the field separator character for pedigree file
`sep.phe`	the field separator character for phenotype file
`sep.gen`	the field separator character for genotype file

Details

Similar to the details for `lmepack.batch.imputed`.

Value

No value is returned. Instead, results are written to `outfile`.

`phen`	phenotype name
`snp`	SNP name
`N`	the number of individuals in analysis
`AF`	imputed allele frequency of coded allele
`h2q`	the portion of phenotypic variation explained by the SNP
`beta`	regression coefficient of SNP covariate
`se`	standard error of `beta`
`pval`	p-value of testing `beta` not equal to zero

Author(s)

Qiong Yang <qyang@bu.edu> and Ming-Huei Chen <mhchen@bu.edu>

References

coxme package: mixed-effects Cox models, sparse matrices, and modeling data from large pedigrees. Beth Atkinson (atkinson@mayo.edu) for pedigree functions.Terry Therneau (therneau@mayo.edu) for all other functions. 2007. Ref Type: Computer Program http://cran.r-project.org/.

Abecasis, G. R., Cardon, L. R., Cookson, W. O., Sham, P. C., & Cherny, S. S. Association analysis in a variance components framework. *Genet Epidemiol*, **21** Suppl 1, S341-S346 (2001).

Examples

```
## Not run:
lmeVpack.batch.imputed(phenfile="simphen.csv",genfile="simgen.csv",pedfile="simped.csv",
phen="SIMQT",kinmat="simkmat.Rdata",outfile="simout.csv",covars=c("age","sex"),
sep.ped=",",sep.phe=",",sep.gen=",")

## End(Not run)
```

---

qq *function to make Qantile-Qantile (QQ) plot for p-values*

Description

qq function makes the QQ plot of p-values against a uniform (0,1) distribution. The genomic control parameter for one degree freedom chi-square statistics corresponding to the p-values is also plotted.

Usage

```
qq(pvalue, outfile)
```

Arguments

pvalue	P-values of interest.
outfile	the file name (xxxx.bmp) for output QQ plot in bitmap format

Author(s)

Qiong Yang <qyang@bu.edu> and Ming-Huei Chen <mhchen@bu.edu>

Examples

```
## Not run:
qq(pvalue=result[,"pval"],outfile="QQ_test.bmp")

## End(Not run)
```

Index